COVID-19 Case Surveillance Public Use Data Utility Summary

Users should consider the level of completeness, including suppression levels when planning their analyses and use of public datasets. Privacy protections will suppress field values to reduce reidentification risks. Completeness varies by jurisdiction (i.e., state, local, and territorial) and time period. Variables are consistently coded to the value "Unknown" when jurisdictions specify in the case data submitted to CDC that the value is unknown, the value "Missing" when jurisdictions do not provide a value, and the value "NA" when the value is suppressed as part of privacy protections.

Dataset version: 2021-08-03

Total records in dataset: 27,879,433

Quick Summary

	all_fields	quasi_fields
total_fields	12	3
total_records	27,879,433	27,879,433
total_cells	334,553,196	83,638,299
missing_fields	67,908,512	896,771
missing_pct	20%	1%
complete_fields	266,644,684	82,741,528
complete_pct	80%	99%
unknown_fields	22,637,875	9,702,091
unknown_pct	7%	12%
suppressed_fields	115	115
suppressed_pct	0%	0%
available_fields	244,006,694	73,039,322
available_pct	73%	87%

Field-level Utility Summary

	suppressed	suppressed_percent	missing	missing_percent
race_ethnicity_combined	16	0.0%	643,821	2.3%
sex	25	0.0%	55,437	0.2%
age_group	74	0.0%	197,513	0.7%
records_with_any_field	74	0.0%	836,565	3.0%